Response to Knoppers et al.

We appreciate the interest by Knoppers et al. in the issues raised by our study.[1] We agree with the authors of this letter on anonymity as a "valid and vital concept" that "may need to be reconceptualized to remain relevant in 21st century science and medicine." We share the concerns of the authors on a nuanced treatment of the subject, but they make some points on which we would elaborate. Knoppers et al. argue that "reidentification risks are minimal" in genome-wide association studies that are presently being conducted. We note that although the theoretical probabilities of reidentification and phenotype reconstruction are real[2–5] and increase, in some cases quite substantially, with the number of phenotypes included in a study (as, for example, with the present use of -omics technologies),[1] it nevertheless remains unclear that malicious attempts to reidentify subjects and/or to reconstruct their phenotypes will ever be anything but exceedingly rare. We recognize that despite their rarity, reidentification events may have unanticipated consequences (e.g., for the study participants and, possibly, for the public perception of genomics research). All things considered, however, we agree that the overall risk of reidentification is likely to be minimal. Moreover, any harm associated with reidentification would be quite variable depending on the phenotypes included in the study, and factors such as the ages and temperaments of participants, and should in many situations be minimal as well. Knoppers et al. also state that "models currently exist that facilitate dissemination of useful health data without compromising privacy," and we agree that dbGaP policies do indeed facilitate data dissemination without compromising study participant privacy.[6] But current policies at dbGaP (and other data repositories) hold dissemination of results of genome investigations to the same standards as dissemination of individual-level data. These restrictive policies reverse long-standing traditions that complete results of published studies should be readily available to the scientific community for inspection, and we believe that they genuinely impede research progress as well. The harm to society (including study participants) associated with slowing, and perhaps sometimes preventing, research discoveries because results of expensive, taxpayer-funded studies are not broadly disseminated should also be considered in determining policies for dissemination of results.

Our conviction is that results of genome studies should always be broadly disseminated unless there are very good reasons to require more restricted dissemination. At the very least, results from large-scale genome studies should be easily accessible to investigators who certify on a regular basis that they will make no attempt to reidentify study participants. We have always served results of our genome studies not involving human subjects (studies on cell lines in the public arena) in public-access databases (e.g., http://www.scandb.org and http://www.PACdb.org) and are now transmitting complete results of our genome studies on human subjects to dbGaP in hopes that policies there will soon permit broad and less restrictive dissemination of those results to the scientific community. We hope that others will join us.

Hae Kyung Im,[1,*] Eric R. Gamazon,[2] Dan L. Nicolae,[2,3,4] and Nancy J. Cox[2,3,*]
[1]Department of Health Studies, [2]Department of Medicine, [3]Department of Human Genetics, [4]Department of Statistics, University of Chicago, Chicago, IL 60637, USA
*Correspondence: haky@uchicago.edu (H.K.I.), ncox@bsd.uchicago.edu (N.J.C.)

References

1. Im, H.K., Gamazon, E.R., Nicolae, D.L., and Cox, N.J. (2012). On sharing quantitative trait GWAS results in an era of multiple-omics data and the limits of genomic privacy. Am. J. Hum. Genet. *90*, 591–598.
2. Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. *4*, e1000167.
3. Jacobs, K.B., Yeager, M., Wacholder, S., Craig, D., Kraft, P., Hunter, D.J., Paschal, J., Manolio, T.A., Tucker, M., Hoover, R.N., et al. (2009). A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. Nat. Genet. *41*, 1253–1257.
4. Visscher, P.M., and Hill, W.G. (2009). The limits of individual identification from sample allele frequencies: theory and statistical analysis. PLoS Genet. *5*, e1000628.
5. Sankararaman, S., Obozinski, G., Jordan, M.I., and Halperin, E. (2009). Genomic privacy and limits of individual detection in a pool. Nat. Genet. *41*, 965–967.
6. Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. Nat. Genet. *39*, 1181–1186.

http://dx.doi.org/10.1016/j.ajhg.2012.07.025.